

Learning Coefficient in Bayesian Estimation of Restricted Boltzmann Machine

Miki Aoyagi¹

¹ *Department of Mathematics, College of Science & Technology, Nihon University, 1-8-14, Surugadai, Kanda, Chiyoda-ku, Tokyo 101-8308, Japan*

Abstract. We consider the real log canonical threshold for the learning model in Bayesian estimation. This threshold corresponds to a learning coefficient of generalization error in Bayesian estimation, which serves to measure learning efficiency in hierarchical learning models [30, 31, 33]. In this paper, we clarify the ideal which gives the log canonical threshold of the restricted Boltzmann machine and consider the learning coefficients of this model.

2000 Mathematics Subject Classifications: 62D05, 62M20, 32S10, 14Q15

Key Words and Phrases: Learning coefficient, resolution of singularities, generalization error, training error

1. Introduction

Many recent studies in statistics have used algebraic geometry, one example being the study of learning coefficients in singular Bayesian statistics. The learning coefficient corresponds to the main term of the generalization and training errors. Singular models, such as neural networks, normal mixtures, reduced rank regressions, Bayesian networks, binomial mixtures, Boltzmann machines, and hidden Markov models, have a singular Fisher metric which is not always approximated by any quadratic form. Therefore, the classic model selection methods of regular statistical models such as AIC [1], TIC [29], HQ [15], NIC [21], BIC [28], and MDL [25], cannot apply to the singular models and it is difficult to analyze the asymptotic behavior of these generalization and training errors.

S. Watanabe [30, 31, 32, 33] established the theory of learning coefficients using algebraic geometry, even though the statistical model is singular. However, he presents only a universal law, and more mathematical consideration is needed to obtain the exact learning coefficient.

The learning coefficient is the log canonical threshold of the Kullback function. Let f be a nonzero holomorphic function over \mathbb{C} or an analytic function over \mathbb{R} on a smooth

Email addresses: aoyagi.miki@math.cst.nihon-u.ac.jp (M. Aoyagi)

variety Y , and let $Z \subset Y$ a closed subscheme. The log canonical threshold $\lambda_Z(Y, f)$ is analytically defined as

$$\lambda_Z(Y, f) = \sup\{c : |f|^{-c} \text{ is locally } L^2 \text{ near } Z\}$$

over \mathbb{C} , and

$$\lambda_Z(Y, f) = \sup\{c : |f|^{-c} \text{ is locally } L^1 \text{ near } Z\}$$

over \mathbb{R} [18, 22]. It is known that if f is a polynomial or a convergent power series, $\lambda_0(\mathbb{C}^d, f)$ is the largest root of the Bernstein-Sato polynomial $b(s) \in \mathbb{C}[s]$ of f , where $b(s)f^s = Pf^{s+1}$ for a linear differential operator P [9, 10, 17]. The log canonical threshold $\lambda_Z(Y, f)$ also corresponds to the largest pole of $\int_{\text{near } Z} |f|^{2z}\psi(w)dw$ over \mathbb{C} , ($\int_{\text{near } Z} |f|^z\psi(w)dw$ over \mathbb{R}), where $\psi(w)$ is a C^∞ -function with a compact support and such that $\psi(w) \neq 0$ on Z .

Log canonical thresholds can be obtained by Hironaka's Theorem [16]. However, it is still difficult to obtain them in learning theory for several reasons, such as degeneration with respect to their Newton polyhedra and non-isolation of their singularities [14]. Moreover, in algebraic geometry and algebraic analysis, these studies are usually done over an algebraically closed field [18, 22]. There are many differences between the real field and the complex field. For example, log canonical thresholds over the complex field are less than 1, while those over the real field are not necessarily less than 1. We cannot therefore apply results over an algebraically closed field to our cases directly.

In this paper, we clarify the ideal which gives the log canonical threshold of the restricted Boltzmann machine (Theorem 1) and consider the learning coefficients of this model (Theorem 2). We use inclusion of ideals and a recursive blowing up from algebraic geometry. We obtain the exact values in certain conditions and bounds in all other cases.

In the past few years, we have also obtained the learning coefficients for reduced rank regression [8] and for Vandermonde matrix type singularities (e.g. the three layered neural network) [7, 2, 6]. Learning coefficients in the case of the normal mixture models with dimension one have also been obtained recently [4]. D. Rusakov, D. Geiger [26, 27] and P. Zwiernik [35] obtained them for naive Bayesian networks and for directed tree models with hidden variables, respectively. In the previous paper [5], we considered the learning coefficient of a modified and simplified restricted Boltzmann machine. The several techniques in [5] are used in this paper (Theorem 3 and Lemma 4).

This paper consists of four sections. In Section 2, we summarize the framework of Bayesian learning models. In Section 3, we show our main results, and we discuss with a relation between our results and other works in Section 4.

2. Learning coefficients and singular fluctuations

In this section, we present the theory of learning coefficients and singular fluctuations.

Let $q(x)$ be a true probability density function of variables $x \in \mathbb{R}^N$ and let $x^n := \{x_i\}_{i=1}^n$ be n training samples selected from $q(x)$, independently and identically. Consider a learning model which is written in probabilistic form as $p(x|w)$, where $w \in W \subset \mathbb{R}^d$ is

a parameter. The purpose of the learning system is to estimate $q(x)$ from x^n by using $p(x|w)$.

Let $p(w|x^n)$ be the *a posteriori* probability density function

$$p(w|x^n) = \frac{1}{Z_n} \psi(w) \prod_{i=1}^n p(x_i|w),$$

where $\psi(w)$ is an *a priori* probability density function on the parameter set W and

$$Z_n = \int_W \psi(w) \prod_{i=1}^n p(x_i|w) dw.$$

Let us define

$$E_w[f(w)] = \frac{\int dw f(w) \psi(w) \prod_{i=1}^n p(x_i|w)^\beta}{\int dw \psi(w) \prod_{i=1}^n p(x_i|w)^\beta},$$

where β is the inverse temperature. We usually set $\beta = 1$.

Then we have a predictive density function $p(x|X^n) = E_w[p(x|w)]$, which is the average inference of the Bayesian density function.

We next introduce the Kullback function $K(q||p)$ and the empirical Kullback function $K_n(q||p)$:

$$K(q||p) = \sum_x q(x) \log \frac{q(x)}{p(x)},$$

$$K_n(q||p) = \frac{1}{n} \sum_{i=1}^n \log \frac{q(x_i)}{p(x_i)},$$

for density functions $p(x), q(x)$. The function $K(p||q)$ always has a positive value and satisfies $K(q||p) = 0$ if and only if $q(x) = p(x)$.

Now we define the Bayesian generalization error B_g , Bayesian training error B_t , Gibbs generalization error G_g and Gibbs training error G_t as follows:

$$B_g = K(q(x)||E_w[p(x|w)]),$$

$$B_t = K_n(q(x)||E_w[p(x_i|w)]),$$

$$G_g = E_w[K(q(x)||p(x|w))],$$

and

$$G_t = E_w[K_n(q(x)||p(x|w))].$$

The most important of these is the Bayesian generalization error. This error shows how precisely the predictive function approximates the true density function.

S. Watanabe [30, 31, 34] proved the following four relations:

$$E[B_g] = \frac{\lambda + \nu\beta - \nu}{n\beta} + o\left(\frac{1}{n}\right)$$

$$E[B_t] = \frac{\lambda - \nu\beta - \nu}{n\beta} + o\left(\frac{1}{n}\right)$$

$$E[G_g] = \frac{\lambda + \nu\beta}{n\beta} + o\left(\frac{1}{n}\right)$$

$$E[G_t] = \frac{\lambda - \nu\beta}{n\beta} + o\left(\frac{1}{n}\right).$$

λ is called a learning coefficient and ν a singular fluctuation, both of which are birational invariants. Mathematically, λ is equal to the log canonical threshold introduced in Section 1. For regular models, it holds that $\lambda = \nu = d/2$ where d is the dimension of the parameter space.

Thus we have

$$E[B_g] = E[B_t] + 2\beta(E[G_t] - E[B_t]) + o\left(\frac{1}{n}\right)$$

and

$$E[G_g] = E[G_t] + 2\beta(E[G_t] - E[B_t]) + o\left(\frac{1}{n}\right).$$

Eliminating the expectation of the true probability density function from these four errors, let

$$BL_g = -\sum_x q(x) \log E_w[p(x|w)],$$

$$BL_t = -\frac{1}{n} \sum_{i=1}^n \log E_w[p(x_i|w)],$$

$$GL_g = -E_w\left[\sum_x q(x) \log p(x|w)\right]$$

$$GL_t = -E_w\left[\frac{1}{n} \sum_{i=1}^n \log p(x_i|w)\right].$$

Then we have

$$E[BL_g] = E[BL_t] + 2\beta(E[G_t] - E[B_t]) + o\left(\frac{1}{n}\right)$$

and

$$E[GL_g] = E[GL_t] + 2\beta(E[G_t] - E[B_t]) + o\left(\frac{1}{n}\right).$$

This is called WAIC (widely applicable information criterion). These two equations show that we can estimate the Bayesian and Gibbs generalization errors from the Bayesian and Gibbs training errors without any knowledge of the true probability density functions. Note that the generalization errors relate to the generalization losses via the entropy of the true distribution. Training errors are calculated from training samples x_i using a learning model p . In real applications or experiments, we usually do not know the true distribution but only the value of the training errors. Our purpose is to estimate the true distribution from the training samples, which shows that these relations are effective.

We can select a suitable model from among several statistical models by observing these values.

The difference between the Bayesian training error and Gibbs training error converges to ν/n :

$$n\beta(E[G_t] - E[B_t]) \rightarrow \nu, \quad n \rightarrow \infty.$$

These relations were shown using resolution of singularities and the Schwarz distribution. Define an empirical process

$$\xi_n(w) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{K(w) - \log \frac{q(x_i)}{p(x_i|w)}}{\sqrt{K(w)}},$$

where $K(w) = \sum_x q(x) \log \frac{q(x)}{p(x|w)}$. If $K(w) = 0$, ξ_n is ill-defined. Let us consider a manifold by a resolution map μ of singularities (Appendix A.2). On a small open set U_α with a local coordinate system $u_\alpha = (u_1, \dots, u_d)$, we have

$$K(\mu(u_\alpha)) = u_1^{2s_1} u_2^{2s_2} \dots u_d^{2s_d},$$

$$\log \frac{q(x)}{p(x|\mu(u_\alpha))} = a(x, u_\alpha) u_1^{s_1} u_2^{s_2} \dots u_d^{s_d},$$

and

$$\psi(\mu(u_\alpha))|\mu'(u_\alpha)| = \psi_0(u_\alpha) u_1^{k_1} u_2^{k_2} \dots u_d^{k_d},$$

where $a(x, u_\alpha)$ satisfies $\sum_x a(x, u_\alpha) q(x) = u_1^{s_1} u_2^{s_2} \dots u_d^{s_d}$ and $\psi_0(u_\alpha) \neq 0$. The value λ is obtained by $\min_\alpha \min_{1 \leq j \leq d} \frac{k_j+1}{2s_j}$, and θ by the max number of elements in $\{j \mid \frac{k_j+1}{2s_j} = \lambda\}$. Also we have

$$\begin{aligned} \xi_n(\mu(u_\alpha)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{u_1^{2s_1} u_2^{2s_2} \dots u_d^{2s_d} - a(x_i, u_\alpha) u_1^{s_1} u_2^{s_2} \dots u_d^{s_d}}{u_1^{s_1} u_2^{s_2} \dots u_d^{s_d}} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (u_1^{s_1} u_2^{s_2} \dots u_d^{s_d} - a(x_i, u_\alpha)), \end{aligned}$$

which is a well-defined empirical process even if $K(w) = 0$. The variable $\xi_n(\mu(u_\alpha))$ converges to a random variable of a Gaussian process $\xi(u_\alpha)$ whose mean is 0 and variance is 2.

By using Schwarz distribution, the value ν is obtained theoretically from the learning coefficient λ and its order θ :

$$\nu = \frac{1}{2} E_\xi \frac{\int_0^\infty dt \sum_{u^*} \int du \xi(u) t^{\lambda-1/2} e^{-\beta t + \beta \sqrt{t} \xi(u)}}{\int_0^\infty dt \sum_{u^*} \int du t^{\lambda-1/2} e^{-\beta t + \beta \sqrt{t} \xi(u)}},$$

where \sum_{u^*} shows the sum of local coordinates that attain the minimum λ and the maximum θ . Our purpose in this paper is to obtain λ for the restricted Boltzmann machine.

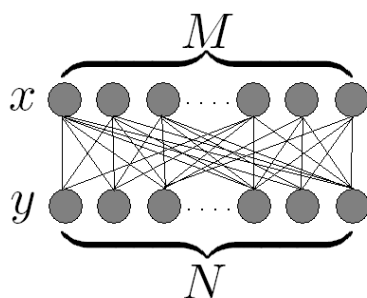


Figure 1: A restricted Boltzmann machine: M is the number of binary observable units x and N is the number of binary hidden units y . The learning model is $p(x, y|a, b, c) \propto \exp(\sum_{i=1}^M \sum_{j=1}^N a_{ij}x_iy_j + \sum_{i=1}^M b_ix_i + \sum_{j=1}^N c_jy_j)$, where a, b, c are parameters.

3. Restricted Boltzmann machine

Definition 1. For a real analytic function f in a neighborhood of w^* , let $\lambda_{w^*}(f)$ be the largest pole of $\int_U |f|^z dw$ and $\theta_{w^*}(f)$ be its order, where U is a sufficiently small neighborhood of w^* . For the ideal I generated by real analytic functions f_1, \dots, f_n in a neighborhood of w^* , define $\lambda_{w^*}(I) = \lambda_{w^*}(f_1^2 + \dots + f_n^2)$.

The second definition is well-defined by Lemma 1 in Appendix A.1.

Let $2 \leq M \in \mathbb{N}$ and $N \in \mathbb{N}$. Set

$$p(x, y|a, b, c) = \frac{\exp(\sum_{i=1}^M \sum_{j=1}^N a_{ij}x_iy_j + \sum_{i=1}^M b_ix_i + \sum_{j=1}^N c_jy_j)}{Z(a, b, c)},$$

where $a \in \mathbb{R}^{M \times N}$, $b \in \mathbb{R}^M$, $c \in \mathbb{R}^N$ are parameters,

$$Z(a, b, c) = \sum_{x_i=0,1, y_j=0,1} \exp(\sum_{i=1}^M \sum_{j=1}^N a_{ij}x_iy_j + \sum_{i=1}^M b_ix_i + \sum_{j=1}^N c_jy_j),$$

$x = (x_i) \in \{0, 1\}^M$ and $y = (y_j) \in \{0, 1\}^N$ (Fig. 1).

Consider a restricted Boltzmann machine

$$\begin{aligned} p(x|a, b, c) &= \sum_{y_j=0,1} p(x, y|a, b, c) \\ &= \sum_{y_j=0,1} \frac{\exp(\sum_{i=1}^M \sum_{j=1}^N a_{ij}x_iy_j + \sum_{i=1}^M b_ix_i + \sum_{j=1}^N c_jy_j)}{Z(a, b, c)} \\ &= \sum_{y_j=0,1} \frac{\prod_{i=1}^M \exp(b_ix_i) \prod_{j=1}^N \exp((\sum_{i=1}^M a_{ij}x_i + c_j)y_j)}{Z(a, b, c)} \\ &= \frac{\prod_{i=1}^M \exp(b_ix_i) \prod_{j=1}^N (1 + \exp(\sum_{i=1}^M a_{ij}x_i + c_j))}{Z(a, b, c)}. \end{aligned}$$

Assume that the true distribution is $p(x|a^*, b^*, c^*)$ with $a^* = (a_{ij}^*), b^* = (b_i^*), c^* = (c_j^*)$. Then the Kullback function $K(a, b, c)$ is

$$\sum_{x_i=0,1} p(x|a^*, b^*, c^*) (\log p(x|a^*, b^*, c^*) - \log p(x|a, b, c)).$$

Since we consider a neighborhood of $\frac{p(x|a,b,c)}{p(x|a^*,b^*,c^*)} = 1$, we have

$$\begin{aligned} & \sum_{x_i=0,1} p(x|a^*, b^*, c^*) (\log p(x|a^*, b^*, c^*) - \log p(x|a, b, c)) \\ = & \sum_{x_i=0,1} p(x|a^*, b^*, c^*) \left(1 - \frac{p(x|a, b, c)}{p(x|a^*, b^*, c^*)} + \frac{1}{2} \left(1 - \frac{p(x|a, b, c)}{p(x|a^*, b^*, c^*)}\right)^2 \{1 + o(1 - \frac{p(x|a, b, c)}{p(x|a^*, b^*, c^*)})\}\right) \\ = & \sum_{x_i=0,1} \left\{ p(x|a^*, b^*, c^*) - p(x|a, b, c) + \frac{p(x|a^*, b^*, c^*)}{2} \left(1 - \frac{p(x|a, b, c)}{p(x|a^*, b^*, c^*)}\right)^2 \{1 + o(1 - \frac{p(x|a, b, c)}{p(x|a^*, b^*, c^*)})\} \right\} \\ = & \sum_{x_i=0,1} \frac{p(x|a^*, b^*, c^*)}{2} \left(1 - \frac{p(x|a, b, c)}{p(x|a^*, b^*, c^*)}\right)^2 \{1 + o(1 - \frac{p(x|a, b, c)}{p(x|a^*, b^*, c^*)})\}. \end{aligned}$$

Therefore we only need to obtain the log canonical threshold of

$$\Psi_0 = \sum_{x_i=\pm 1} (p(x|a, b, c) - p(x|a^*, b^*, c^*))^2,$$

since the ideal generated by $1 - \frac{p(x|a,b,c)}{p(x|a^*,b^*,c^*)}$, $x \in \{0, 1\}^M$ is equal to the ideal generated by $p(x|a, b, c) - p(x|a^*, b^*, c^*)$, $x \in \{0, 1\}^M$ (Definition 1).

Set $\mathcal{I} = \{x = (x_i) \in \{0, 1\}^M\}$ and write $|x| = x_1 + \dots + x_M$.

By Theorem 3 in Appendix A.1, rather than considering the function Ψ_0 , we only need to obtain the log canonical threshold of

$$\begin{aligned} & \sum_{x \in \mathcal{I}} \left\{ \sum_{i=1}^M (b_i x_i) + \sum_{j=1}^N \log(1 + \exp(\sum_{i=1}^M a_{ij} x_i + c_j)) - \sum_{i=1}^M (b_i^* x_i) - \sum_{j=1}^N \log(1 + \exp(\sum_{i=1}^M a_{ij}^* x_i + c_j^*)) \right. \\ & \left. - \sum_{j=1}^N \log(1 + \exp(c_j)) + \sum_{j=1}^N \log(1 + \exp(c_j^*)) \right\}^2, \end{aligned} \tag{1}$$

where $x_0 = (0, \dots, 0)$.

Theorem 1. *Let*

$$\begin{aligned} b'_i &= b_i - b_i^* + \sum_{j=1}^N \log\left(\frac{1 + \exp(a_{ij} + c_j)}{1 + \exp(\sum_{i=1}^M a_{ij}^* + c_j^*)}\right) - \sum_{j=1}^N \log\left(\frac{1 + \exp c_j}{1 + \exp c_j^*}\right), \\ a'_{ij} &= \frac{1 - \exp(a_{ij})}{1 + \exp(a_{ij}) \exp(c_j)}, \text{ and } c'_j = \exp(c_j). \end{aligned}$$

Also let

$$s_{x,j,m} = \sum_{\substack{i_1 < i_2 < \dots < i_m \\ x_{i_1} = x_{i_2} = \dots = x_{i_m} = 1}} a'_{i_1 j} a'_{i_2 j} \dots a'_{i_m j}.$$

The log canonical threshold of the restricted Boltzmann machine is equal to that of

$$\sum_{i=1}^M b_i'^2 + \sum_{x \in \mathcal{I}, |x| \geq 2} \left\{ \sum_{j=1}^N \log \left(\frac{\sum_{m=2}^{|x|} \{c_j'(c_j'^{m-2} - c_j'^{m-3} + c_j'^{m-4} - \dots + (-1)^m\} s_{x,j,m} - c_j^{*'}(c_j^{*'}{}^{m-2} - \dots + (-1)^m) s_{x,j,m}^*}{\sum_{m=2}^{|x|} c_j^{*'}(c_j^{*'}{}^{m-2} - c_j^{*'}{}^{m-3} + c_j^{*'}{}^{m-4} - \dots + (-1)^m) s_{x,j,m}^* + 1} + 1 \right) \right\}^2, \tag{2}$$

where $c_j' > 0$ for $j = 1, \dots, N$.

The proof of this theorem appears in Appendix A.1.

Note that if the above theorem is not used, the function $Z(a, b, c)$ is needed to be clarified in more detail for our purpose.

Theorem 2. Consider the restricted Boltzmann machine

$$p(x|a, b, c) = \frac{\prod_{i=1}^M \exp(b_i x_i) \prod_{j=1}^N (1 + \exp(\sum_{i=1}^M a_{ij} x_i + c_j))}{Z(a, b, c)},$$

where a, b, c are parameters and the true distribution is $p(x|a^*, b^*, c^*)$.

Case 1 If $M > N$ and $a^* = 0$, then we have $\lambda = \frac{M}{2} + \frac{MN}{4}$.

Case 2 If $M \leq N$ and $a^* = 0$, then we have $\frac{M}{2} + \frac{M(M-1)}{4} \leq \lambda \leq \frac{M}{2} + \frac{MN}{4}$.

Case 3 We have $\lambda \leq \frac{M}{2} + \sum_{j=1}^N \lambda_j$, where

$$\lambda_j = \begin{cases} \frac{3M}{4}, & r_j = 0, \\ \frac{2M-1}{2}, & r_j = 1, 2, \\ \frac{2M+1}{2}, & r_j \geq 3, \end{cases}$$

where r_j is the number of elements in $\{a_{ij}^* \neq 0, i = 1, \dots, M\}$.

Case 4 If $N = 1$ then $\lambda = \begin{cases} \frac{3M}{4}, & r = 0, \\ \frac{2M-1}{2}, & r = 1, 2, \\ \frac{2M+1}{2}, & r \geq 3. \end{cases}$ and $\theta = \begin{cases} 2, & M = 2, r = 0, \\ 1, & \text{otherwise,} \end{cases}$ where r

is the number of elements in $\{a_{i1}^* \neq 0, i = 1, \dots, M\}$.

Case 5 If $M = 2$ then $\lambda = \frac{3}{2}$ and $\theta = \begin{cases} 2, & N = 1, a^* = 0, \\ 1, & \text{otherwise.} \end{cases}$

The proof of this theorem appears in Appendix A.2.

Remark 1 When the true distribution is not in the restricted Boltzmann machine model, the learning coefficient λ is infinity, because the Kullback function is positive.

Remark 2 Rusakov & Geiger [26, 27] obtained λ and θ for the following class of Naive Bayesian networks with two hidden states and binary features:

$$p(x|d, e, t) = t \prod_{i=1}^M d_i^{x_i} (1 - d_i)^{1-x_i} + (1 - t) \prod_{i=1}^M e_i^{x_i} (1 - e_i)^{1-x_i},$$

where $x \in \{0, 1\}^M$, $d = \{d_i\}_{i=1}^M \in (0, 1)^M$, $e = \{e_i\}_{i=1}^M \in (0, 1)^M$ and $0 \leq t \leq 1$. For symmetry, we can assume that $0 < t \leq 1$. Set $\exp(b_i) = \frac{d_i}{1-d_i}$ and $\exp(a_{i1} + b_i) = \frac{e_i}{1-e_i}$.

If we set $\frac{1-t}{t} \prod_{i=1}^N \frac{1-e_i}{1-d_i} = c'_1$, then we have

$$p(x|d, e, t) = \frac{\prod_{i=1}^N \exp(b_i x_i) + c'_1 \prod_{i=1}^N \exp((a_{i1} + b_i) x_i)}{Z(a, b, c)},$$

where $Z(a, b, c) = \frac{1}{t \prod_{i=1}^N (1-d_i)}$.

The difference between two models is as follows: $c'_1 \geq 0$ for Naive Bayesian networks, while $c'_1 = \exp c_1 > 0$ for the restricted Boltzmann machine.

$$\text{We have } \lambda = \begin{cases} \frac{M+1}{2}, & c'_1 = 0, \\ \frac{3M}{4}, & c'_1 \neq 0, r = 0, \\ \frac{2M-1}{2}, & c'_1 \neq 0, r = 1, 2, \\ \frac{2M+1}{2}, & c'_1 \neq 0, r \geq 3. \end{cases} \quad \text{and } \theta = \begin{cases} 3, & c'_1 = 0, M = 2, r = 0, \\ 2, & c'_1 \neq 0, M = 2, r = 0, \\ 1, & \text{otherwise,} \end{cases}$$

where r is the number of elements in $\{a_{i1}^* \neq 0, i = 1, \dots, M\}$ [26, 27]. (The values λ and θ in the case $c'_1 \neq 0, r = 0$ are added to those in [26, 27].)

A shorter proof than that in [26, 27] appears in Appendix A.2.4 and A.2.6. This seems to suggest that Theorem 1 is effective.

4. Discussion

In this paper, we consider the learning coefficient of restricted Boltzmann machines (Theorem 2). We obtain the exact values for $M > N$, $a^* = 0$, for $M = 2$ and for $N = 1$, and give bounds for the other types. The case $M > N$ is usually the case in applications. Case 4 shows that λ is not linear with respect to r and case 5 shows that λ is independent of the number of elements $a_{ij}^* \neq 0$. These observations imply that we need more careful consideration to obtain the exact values λ when $a^* \neq 0$.

The learning coefficients of our recent results have been used very effectively by M. Drton [12, 13] for model selection, using a method called sBIC, which is available for singular models, while BIC cannot be applied to singular models.

Our theoretical results introduce a mathematical measure of precision to numerical calculations such as Markov Chain Monte Carlo. Nagata and Watanabe [23, 24] gave a

mathematical foundation for analyzing and developing the precision of the MCMC method using our theoretical values of marginal likelihoods.

G. Montufar and N. Ay [20] showed a universal approximation result for the restricted Boltzmann machine. They showed that any distribution on $\{0, 1\}^M$ can be approximated arbitrarily well by $N = 2^{M-1} - 1$ hidden units. By using Theorem 3 and by setting

$$b'_i = b_i + \sum_{j=1}^N \log(1 + \exp(a_{ij} + c_j)) - \sum_{j=1}^N \log(1 + \exp c_j),$$

$$a'_{ij} = \frac{1 - \exp(a_{ij})}{1 + \exp(a_{ij}) \exp(c_j)}, \text{ and } c'_j = \exp(c_j),$$

and

$$s_{x,j,m} = \sum_{\substack{i_1 < i_2 < \dots < i_m \\ x_{i_1} = x_{i_2} = \dots = x_{i_m} = 1}} a'_{i_1 j} a'_{i_2 j} \dots a'_{i_m j},$$

their theorem showed that if $N \geq 2^{M-1} - 1$, then there exist $b_i^{*'}, a_{ij}^{*'}, c_j^{*'}$ such that $b_i^{*'} = \log \frac{\tilde{W}(\mathbf{e}_i)}{\tilde{W}(\mathbf{0})}$, $\sum_{j=1}^N \log\{1 + \sum_{m=2}^{|x|} c_j^{*'} (c_j^{*'}{}^m - c_j^{*'}{}^{m-1} + \dots + (-1)^m) s_{x,j,m}\} = \log \frac{\tilde{W}(x)}{\tilde{W}(\mathbf{0})}$ for any distribution $\tilde{p}(x) = \frac{\tilde{W}(x)}{\sum_{x \in \{0,1\}^M} \tilde{W}(x)} > 0$ on $x \in \{0, 1\}^M$, where $\mathbf{e}_i \in \{0, 1\}^M$ such that i th element of \mathbf{e}_i is 1 and the other elements are 0. Then we have $\lambda_{b_i^{*'}, a_{ij}^{*'}, c_j^{*'}}(J) \leq \min\{\frac{2^n - 1}{2}, \frac{MN + N + M}{2}\}$ where

$$J = \left\langle b'_i - \log \frac{\tilde{W}(\mathbf{e}_i)}{\tilde{W}(\mathbf{0})}, \sum_{j=1}^N \log\{1 + \sum_{m=2}^{|x|} c'_j (c_j'^m - \dots + (-1)^m) s_{x,j,m}\} - \log \frac{\tilde{W}(x)}{\tilde{W}(\mathbf{0})} \right\rangle$$

$$| x \in \{0, 1\}^M, |x| \geq 2 \rangle.$$

Moreover, M. A. Cueto, J. Morton and B. Sturmfels [11] gave the conjecture such that “the restricted Boltzmann machine has the expected dimension $\min\{MN + M + N, 2^M - 1\}$ ” and they proved that it is true when $N \leq 2^{M - \lceil \log_2(M+1) \rceil}$ and $N \geq 2^{M - \lfloor \log_2(M+1) \rfloor}$. This theorem showed that the maximum rank of J 's Jacobian matrix has $\min\{MN + M + N, 2^M - 1\}$.

Roughly speaking, λ shows the character at a singular point, while an approximation result and a result of dimension show the character at general points.

Our future research aims to improve our methods and to apply them for the general Boltzmann machine, which is also known as the Bayesian network model, the graphical model or the spin model, as such models are widely used in many fields. These are new problems, even in mathematics, to obtain the desingularization of Kullback functions, since the singularities of these functions are very complicated and as such most of them have not yet been investigated. However, we believe that extending our results would provide a mathematical foundation for the analysis of various graphical models.

Acknowledgments

This research was supported by the Ministry of Education, Science, Sports and Culture in Japan, Grant-in-Aid for Scientific Research 22540224.

References

- [1] H. Akaike. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, 19:716–723, 1974.
- [2] M. Aoyagi. The zeta function of learning theory and generalization error of three layered neural perceptron. *RIMS Kokyuroku, Recent Topics on Real and Complex Singularities*, 1501:153–167, 2006.
- [3] M. Aoyagi. Log canonical threshold of Vandermonde matrix type singularities and generalization error of a three layered neural network. *International Journal of Pure and Applied Mathematics*, 52(2):177–204, 2009.
- [4] M. Aoyagi. A Bayesian learning coefficient of generalization error and Vandermonde matrix-type singularities. *Communications in Statistics - Theory and Methods*, 39(15):2667–2687, 2010.
- [5] M. Aoyagi. Stochastic complexity and generalization error of a restricted Boltzmann machine in Bayesian estimation. *Journal of Machine Learning Research*, 11(Apr):1243–1272, 2010.
- [6] M. Aoyagi and K. Nagata. Learning coefficient of generalization error in Bayesian estimation and Vandermonde matrix type singularity. *Neural Computation*, 24(6):1569–1610, 2012.
- [7] M. Aoyagi and S. Watanabe. Resolution of singularities and the generalization error with Bayesian estimation for layered neural network. *IEICE Trans. J88-D-II*, 10:2112–2124, 2005a.
- [8] M. Aoyagi and S. Watanabe. Stochastic complexities of reduced rank regression in Bayesian estimation. *Neural Networks*, 18:924–933, 2005b.
- [9] I. N. Bernstein. The analytic continuation of generalized functions with respect to a parameter. *Functional Anal. Appl.*, 6:26–40, 1972.
- [10] J. E. Björk. *Rings of differential operators*. Amsterdam: North-Holland, 1979.
- [11] M. A. Cueto, J. Morton, and B. Sturmfels. Geometry of the restricted Boltzmann machine. *Contemporary Mathematics: Algebraic Methods in Statistics and Probability II*, 516:135–153, 2010.
- [12] M. Drton. Conference lecture: Reduced rank regression. *Workshop on Singular Learning Theory, AIM 2011*, <http://math.berkeley.edu/critch/slt2011/>, 2011.

- [13] M. Drton. Conference lecture: Bayesian information criterion for singular models. *Algebraic Statistics 2012 in the Alleghenies at The Pennsylvania State University*, <http://jasonmorton.com/aspsu2012/>, 2012.
- [14] W. Fulton. *Introduction to toric varieties*, *Annals of Mathematics Studies*. Princeton University Press, 1993.
- [15] E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of Royal Statistical Society Series B*, 41:190–195, 1979.
- [16] H. Hironaka. Resolution of singularities of an algebraic variety over a field of characteristic zero. *Annals of Math*, 79:109–326, 1964.
- [17] M. Kashiwara. B-functions and holonomic systems. *Inventiones Math.*, 38:33–53, 1976.
- [18] J. Kollár. Singularities of pairs. *Algebraic geometry-Santa Cruz 1995, Proc. Sympos. Pure Math., Amsdsr. Math. Soc., Providence, RI*, 62:221–287, 1997.
- [19] S. Lin. Asymptotic approximation of marginal likelihood integrals. (*preprint*), 2010.
- [20] G. Montufar and N. Ay. Refinements of universal approximation results for deep belief networks and restricted Boltzmann machines. *Neural Computation*, 2011, 23(5):1306–1319, 2011.
- [21] N. J. Murata, S. G. Yoshizawa, and S. Amari. Network information criterion - determining the number of hidden units for an artificial neural network model. *IEEE Trans. on Neural Networks*, 5(6):865–872, 1994.
- [22] M. Mustata. Singularities of pairs via jet schemes. *J. Amer. Math. Soc.*, 15:599–615, 2002.
- [23] K. Nagata and S. Watanabe. Exchange Monte Carlo sampling from Bayesian posterior for singular learning machines. *IEEE Transactions on Neural Networks*, 19(7):1253–1266, 2008a.
- [24] K. Nagata and S. Watanabe. Asymptotic behavior of exchange ratio in exchange Monte Carlo method. *International Journal of Neural Networks*, 21(7):980–988, 2008b.
- [25] J. Rissanen. Universal coding, information, prediction, and estimation. *IEEE Trans. on Information Theory*, 30(4):629–636, 1984.
- [26] D. Rusakov and D. Geiger. Asymptotic model selection for naive Bayesian networks. *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, pages 438–445, 2002.
- [27] D. Rusakov and D. Geiger. Asymptotic model selection for naive Bayesian networks. *Journal of Machine Learning Research*, 6:1–35, 2005.

- [28] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.
- [29] K. Takeuchi. Distribution of an information statistic and the criterion for the optimal model. *Mathematical Science*, 153:12–18, 1976.
- [30] S. Watanabe. Algebraic analysis for nonidentifiable learning machines. *Neural Computation*, 13(4):899–933, 2001a.
- [31] S. Watanabe. Algebraic geometrical methods for hierarchical learning machines. *Neural Networks*, 14(8):1049–1060, 2001b.
- [32] S. Watanabe. Algebraic geometry of learning machines with singularities and their prior distributions. *Journal of Japanese Society of Artificial Intelligence*, 16(2):308–315, 2001c.
- [33] S. Watanabe. *Algebraic Geometry and Statistical Learning Theory*, volume 25. Cambridge University Press, 2009.
- [34] S. Watanabe. Equations of states in singular statistical estimation. *Neural Networks*, 23(1):20–34, 2010.
- [35] P. Zwiernik. An asymptotic behavior of the marginal likelihood for general Markov models. *Journal of Machine Learning Research*, 12:3283–3310, 2011.

A. Appendix

A.1. The proof of Theorem 1.

We give below Lemma 1 as it is frequently used in this paper.

Lemma 1 ([3, 4, 19]). *Let U be a neighborhood of $w^* \in \mathbb{R}^d$. Let I be the ideal generated by f_1, \dots, f_n which are analytic functions defined on U .*

- (1) *If $g_1^2 + \dots + g_m^2 \leq f_1^2 + \dots + f_n^2$ then $\lambda_{w^*}(g_1^2 + \dots + g_m^2) \leq \lambda_{w^*}(f_1^2 + \dots + f_n^2)$.*
- (2) *If $g_1, \dots, g_m \in I$, then $\lambda_{w^*}(g_1^2 + \dots + g_m^2) \leq \lambda_{w^*}(f_1^2 + \dots + f_n^2)$. In particular, if g_1, \dots, g_m generate the ideal I then $\lambda_{w^*}(f_1^2 + \dots + f_n^2) = \lambda_{w^*}(g_1^2 + \dots + g_m^2)$.*

Proof.

- (1) If $(g_1^2 + \dots + g_m^2)^{-c}$, $c > 0$ is L^1 then $(f_1^2 + \dots + f_n^2)^{-c}$ is also L^1 .
- (2) The fact $g_1^2 + \dots + g_m^2 \leq P(f_1^2 + \dots + f_n^2)$ for $P \gg 1$ completes the proof.

Q.E.D.

The following lemma is also used in the proofs.

Lemma 2. *Let I, J be the ideals generated by $f_1(w), \dots, f_n(w)$ and $g_1(w'), \dots, g_m(w')$, respectively. If w and w' are different variables then*

$$\lambda_{(w^*, w'^*)}(f_1^2 + \dots + f_n^2 + g_1^2 + \dots + g_m^2) = \lambda_{w^*}(f_1^2 + \dots + f_n^2) + \lambda_{w'^*}(g_1^2 + \dots + g_m^2).$$

The following theorem is a modified version in the previous paper [5].

Theorem 3 ([5]). *Assume that $p(x|a) = \frac{\prod_{j=1}^N W_j(x,a)}{\sum_x \prod_{j=1}^N W_j(x,a)}$ and $q(x) = \frac{\tilde{W}(x)}{\sum_x \prod_{j=1}^N \tilde{W}(x)}$ for $x \in X$. Then the log canonical threshold of $\sum_{x \in X} (p(x|a) - q(x))^2$ and its order are those of*

$$\sum_{x \in X} \left\{ \sum_j^N \log W_j(x, a) - \log \tilde{W}(x) - \sum_j^N \log W_j(x_0, a) + \log \tilde{W}(x_0) \right\}^2,$$

for a fixed $x_0 \in X$.

Proof. Consider the ideal I generated by $p(x|a) - q(x)$ for $x \in X$.

Then I is generated by $\frac{\prod_{j=1}^N W_j(x,a)}{\tilde{W}(x)} - \frac{\sum_x \prod_{j=1}^N W_j(x,a)}{\sum_x \tilde{W}(x)}$, and so by $\frac{\prod_{j=1}^N W_j(x,a)}{\tilde{W}(x)} - \frac{\prod_{j=1}^N W_j(x',a)}{\tilde{W}(x')}$ for $x, x' \in X$.

Since $|x - 1|/2 \leq |\log x| \leq 2|x - 1|$ for $|x - 1| < 1/2$, we have

$$\begin{aligned} & \sum_{x, x' \in X} \left(\frac{\prod_{j=1}^N W_j(x, a)}{\tilde{W}(x)} \frac{\tilde{W}(x')}{\prod_{j=1}^N W_j(x', a)} - 1 \right)^2 / 4 \\ & \leq \sum_{x, x' \in X} \left(\sum_j^N \log W_j(x, a) - \log \tilde{W}(x) + \log \tilde{W}(x') - \sum_j^N \log W_j(x', a) \right)^2 \end{aligned}$$

$$\leq \sum_{x, x' \in X} \left(\frac{\prod_{j=1}^N W_j(x, a)}{\tilde{W}(x)} \frac{\tilde{W}(x')}{\prod_{j=1}^N W_j(x', a)} - 1 \right)^2 4.$$

By Lemma 1 (1), we have the log canonical threshold of $\sum_{x \in X} (p(x|a) - q(x))^2$ and its order are those of the ideal

$$\left\langle \sum_j^N \log W_j(x, a) - \log \tilde{W}(x) + \log \tilde{W}(x') - \sum_j^N \log \tilde{W}(x') \mid x, x' \in X \right\rangle,$$

which is equal to $\left\langle \sum_j^N \log W_j(x, a) - \log \tilde{W}(x) + \log \tilde{W}(x_0) - \sum_j^N \log W_j(x_0, a) \mid x \in X \right\rangle$ for a fixed $x_0 \in X$. The proof is completed by Lemma 1 (2).

Q.E.D.

Set

$$b'_i = b_i - b_i^* + \sum_{j=1}^N \log \left(\frac{1 + \exp(a_{ij} + c_j)}{1 + \exp(\sum_{i=1}^M a_{ij}^* + c_j^*)} \right) - \sum_{j=1}^N \log \left(\frac{1 + \exp c_j}{1 + \exp c_j^*} \right).$$

Then we have

$$\begin{aligned} & \sum_{i=1}^M (b_i - b_i^*) x_i + \sum_{j=1}^N \log \left(\frac{1 + \exp(\sum_{i=1}^M a_{ij} x_i + c_j)}{1 + \exp(\sum_{i=1}^M a_{ij}^* x_i + c_j^*)} \right) - \sum_{j=1}^N \log \left(\frac{1 + \exp(c_j)}{1 + \exp(c_j^*)} \right) \\ &= \sum_{i=1}^M b'_i x_i - \sum_{i=1}^M \sum_{j=1}^N \log \left(\frac{1 + \exp(a_{ij} + c_j)}{1 + \exp(a_{ij}^* + c_j^*)} \right) x_i + \sum_{j=1}^N \log \left(\frac{1 + \exp(\sum_{i=1}^M a_{ij} x_i + c_j)}{1 + \exp(\sum_{i=1}^M a_{ij}^* x_i + c_j^*)} \right) \\ & \quad - \sum_{j=1}^N \log \left(\frac{1 + \exp(c_j)}{1 + \exp(c_j^*)} \right) \left(1 - \sum_{i=1}^M x_i \right), \end{aligned}$$

which is the term in Eq. (1).

Let $a'_{ij} = \frac{1 - \exp(a_{ij})}{1 + \exp(a_{ij}) \exp(c_j)}$ and $c'_j = \exp(c_j)$.

Since

$$\begin{aligned} & \sum_{j=1}^N \log \frac{(1 + c'_j \prod_{i=1}^M \exp(a_{ij} x_i))(1 + c'_j)^{(-1 + \sum_i x_i)}}{\prod_{i=1}^M (1 + \exp(a_{ij}) c'_j)^{x_i}} \\ &= \sum_{j=1}^N \log \frac{\prod_{i=1}^M (1 + a'_{ij} c'_j)^{x_i} (1 + c'_j \prod_{i=1}^M \exp(a_{ij} x_i))(1 + c'_j)^{(-1 + \sum_i x_i)}}{\prod_{i=1}^M (1 + c'_j)^{x_i}} \\ &= \sum_{j=1}^N \log \frac{\prod_{i=1}^M (1 + a'_{ij} c'_j)^{x_i} (1 + c'_j \prod_{i=1}^M \exp(a_{ij} x_i))}{1 + c'_j} \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^N \log \frac{\prod_{i=1}^M (1 + a'_{ij} c'_j)^{x_i} + c'_j \prod_{i=1}^M (1 - a'_{ij})^{x_i}}{1 + c'_j} \\
&= \sum_{j=1}^N \log \left(\frac{\prod_{i=1}^M (1 + a'_{ij} c'_j)^{x_i} + c'_j \prod_{i=1}^M (1 - a'_{ij})^{x_i} - 1 - c'_j}{1 + c'_j} + 1 \right) \\
&= \sum_{j=1}^N \log \left(\frac{\sum_{m=1}^{|x|} c_j'^m s_{x,j,m} + c'_j \sum_{m=1}^{|x|} (-1)^m s_{x,j,m}}{1 + c'_j} + 1 \right) \\
&= \sum_{j=1}^N \log \left(\sum_{m=2}^{|x|} c_j' (c_j'^{m-2} - c_j'^{m-3} + c_j'^{m-4} - \dots + (-1)^m) s_{x,j,m} + 1 \right),
\end{aligned}$$

we have in Eq. (1) as

$$\begin{aligned}
&\sum_{x \in \mathcal{I}} \left\{ \sum_{i=1}^M b'_i x_i + \sum_{j=1}^N \log \left(\sum_{m=2}^{|x|} c_j' (c_j'^{m-2} - c_j'^{m-3} + c_j'^{m-4} - \dots + (-1)^m) s_{x,j,m} + 1 \right) \right. \\
&\left. - \sum_{j=1}^N \log \left(\sum_{m=2}^{|x|} c_j^{*'} (c_j^{*'}{}^{m-2} - c_j^{*'}{}^{m-3} + c_j^{*'}{}^{m-4} - \dots + (-1)^m) s_{x,j,m}^* + 1 \right) \right\}^2, \quad (3)
\end{aligned}$$

where $s_{x,j,m}^* = \sum_{\substack{i_1 < i_2 < \dots < i_m \\ x_{i_1} = x_{i_2} = \dots = x_{i_m} = 1}} a_{i_1 j}^{*'} a_{i_2 j}^{*'} \dots a_{i_m j}^{*'}$.

By Lemma 1, we need to consider the log canonical threshold of

$$\begin{aligned}
&\sum_{i=1}^M b_i'^2 + \sum_{x \in \mathcal{I}, |x| \geq 2} \left\{ \sum_{j=1}^N \log \left(\sum_{m=2}^{|x|} c_j' (c_j'^{m-2} - c_j'^{m-3} + c_j'^{m-4} - \dots + (-1)^m) s_{x,j,m} + 1 \right) \right. \\
&\left. - \sum_{j=1}^N \log \left(\sum_{m=2}^{|x|} c_j^{*'} (c_j^{*'}{}^{m-2} - c_j^{*'}{}^{m-3} + c_j^{*'}{}^{m-4} - \dots + (-1)^m) s_{x,j,m}^* + 1 \right) \right\}^2 \\
&= \sum_{i=1}^M b_i'^2 + \sum_{x \in \mathcal{I}, |x| \geq 2} \left\{ \sum_{j=1}^N \right. \\
&\left. \log \left(\frac{\sum_{m=2}^{|x|} \{ c_j' (c_j'^{m-2} - c_j'^{m-3} + c_j'^{m-4} - \dots + (-1)^m) s_{x,j,m} - c_j^{*'} (c_j^{*'}{}^{m-2} - \dots + (-1)^m) s_{x,j,m}^* \}}{\sum_{m=2}^{|x|} c_j^{*'} (c_j^{*'}{}^{m-2} - c_j^{*'}{}^{m-3} + c_j^{*'}{}^{m-4} - \dots + (-1)^m) s_{x,j,m}^* + 1} + 1 \right) \right\}^2. \quad (4)
\end{aligned}$$

Q.E.D.

A.2. The proof of Theorem 2.

Theorem 4. [Desingularization, Hironaka [16]]

Let f be a real analytic function in a neighborhood of $w^* \in \mathbb{R}^d$ with $f(w^*) = 0$. There exist an open set $V \ni w^*$, a real analytic manifold U , and a proper analytic map μ from U to V such that

- (1) $\mu : U - \mathcal{E} \rightarrow V - f^{-1}(0)$ is an isomorphism, where $\mathcal{E} = \mu^{-1}(f^{-1}(0))$,
- (2) for each $u \in U$, there is a local analytic coordinate system (u_1, \dots, u_d) such that $f(\mu(u)) = \pm u_1^{s_1} u_2^{s_2} \cdots u_d^{s_d}$, where s_1, \dots, s_d are non-negative integers.

Applying Hironaka's theorem to the Kullback function $K(w) \geq 0$, for each $w^* \in K^{-1}(0) \cap W$, we have a proper analytic map μ from an analytic manifold U to a neighborhood V_{w^*} of w^* satisfying Hironaka's Theorem (1) and (2). Then the local integration on V_{w^*} of the zeta function $\int_{V_{w^*}} |K(w)|^z \psi(w) dw$ is

$$\begin{aligned} & \int_{V_{w^*}} |K(w)|^z \psi(w) dw \\ &= \sum_{\alpha} \int_{U_{\alpha}} (u_1^{2s_1} u_2^{2s_2} \cdots u_d^{2s_d})^z \psi(\mu) |\mu'_w(u)| du, \end{aligned}$$

where U_{α} is a small open set with local coordinate system (u_1, \dots, u_d) and $U = \cup U_{\alpha}$. Therefore, the poles can be obtained as $(k_1 + 1)/(2s_1), \dots, (k_d + 1)/(2s_d)$, where $|\mu'_w(u)| = u_1^{k_1} u_2^{k_2} \cdots u_d^{k_d} \mu_0(u)$ and $\mu_0(u) \neq 0$.

It is known that μ in Hironaka's Theorem can be obtained by using a blowing up process.

For simplicity, set $a_{ij} = a'_{ij}$, $b_i = b'_i$, and $c_{ij} = c'_{ij}$.

Let $B^x =$

$$\sum_{j=1}^N \log \left(\frac{\sum_{m=2}^{|x|} \{c_j(c_j^{m-2} - c_j^{m-3} + c_j^{m-4} - \cdots + (-1)^m) s_{x,j,m} - c_j^*(c_j^{*m-2} - \cdots + (-1)^m) s_{x,j,m}^*\}}{\sum_{m=2}^{|x|} c_j^*(c_j^{*m-2} - c_j^{*m-3} + c_j^{*m-4} - \cdots + (-1)^m) s_{x,j,m}^* + 1} + 1 \right).$$

Definition 2. (1) Let $r = (r_{ij})$ be an $H \times H'$ matrix and x an element of $\{0, 1\}^H$. An

analytic function f is called an x -type function of $\begin{pmatrix} r_{i'1} & r_{i'2} & \cdots & r_{i'H'} \\ r_{i'+1,1} & r_{i'+1,2} & \cdots & r_{i'+1,H'} \\ & & \vdots & \\ r_{H1} & r_{H2} & \cdots & r_{HH'} \end{pmatrix}$, if

$$f \in \prod_{x_i=1, i' \leq i \leq H} \langle r_{i1}, \dots, r_{iH'} \rangle.$$

(2) Let $x, x' \in \{0, 1\}^H$. Denote $x \leq x'$ if $x_i \leq x'_i$ for all $i = 1, \dots, H$, and $x < x'$ if $x \leq x'$ and $x \neq x'$.

For example, $s_{x,j,m}$ is an x' -type function of a for all $x' \leq x$ ($x' \in \{0, 1\}^M$).

From now, set $x_{i_1, i_2} = x$ such that $|x| = 2$, $x_{i_1} = x_{i_2} = 1$.

Lemma 3. *We have*

$$\begin{aligned} & \langle B^{x_{i_1, i_2}} : i_1 < i_2 \rangle \subset \langle B^x : |x| \geq 2 \rangle \\ \subset & \left\langle \{c_j(c_j^{m-2} - c_j^{m-3} + \dots + (-1)^m)a_{i_1 j}a_{i_2 j} \cdots a_{i_m j} - c_j^*(c_j^{*m-2} - c_j^{*m-3} + \dots + (-1)^m)a_{i_1 j}^*a_{i_2 j}^* \cdots a_{i_m j}^* \right. \\ & \left. : i_1 < i_2 < \dots < i_m, j = 1, \dots, N\right\rangle. \end{aligned}$$

The following lemma is used for proving Theorem 2.

Lemma 4 ([5]). *After a linear coordinate change from $(r_{i_2}, \dots, r_{i_N})$ to $(r'_{i_2}, \dots, r'_{i_N})$ for $i = 1, \dots, M$ on the manifold obtained by a blowing up with variable (d_2, \dots, d_N) , we have*

$$\begin{aligned} & (r_{i_2}, \dots, r_{i_N}) \begin{pmatrix} d_2 \\ \vdots \\ d_N \end{pmatrix} (d_2, \dots, d_N) \begin{pmatrix} r_{j2} \\ \vdots \\ r_{jN} \end{pmatrix} + (r_{i_2}, \dots, r_{i_N}) \begin{pmatrix} r_{j2} \\ \vdots \\ r_{jN} \end{pmatrix} \\ = & \sum_{k=2}^N r'_{ik} r'_{jk}. \end{aligned}$$

Proof.

Construct the blow-up along the submanifold $\{d_2 = d_3 = \dots = d_N = 0\}$.

Let $d_2 = v_1$, $d_3 = v_1 d'_3$, \dots , $d_N = v_1 d'_N$.

Then we have,

$$\begin{aligned} & (r_{i_2}, \dots, r_{i_N}) \begin{pmatrix} d_2 \\ d_3 \\ \vdots \\ d_N \end{pmatrix} (d_2, \dots, d_N) \begin{pmatrix} r_{j2} \\ r_{j3} \\ \vdots \\ r_{jN} \end{pmatrix} + (r_{i_2}, \dots, r_{i_N}) \begin{pmatrix} r_{j2} \\ r_{j3} \\ \vdots \\ r_{jN} \end{pmatrix} \\ = & (r_{i_2}, \dots, r_{i_N}) \begin{pmatrix} v_1^2 \\ v_1^2 \begin{pmatrix} 1 \\ d'_3 \\ \vdots \\ d'_N \end{pmatrix} \end{pmatrix} (1, d'_3, \dots, d'_N) + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} r_{j2} \\ r_{j3} \\ \vdots \\ r_{jN} \end{pmatrix}. \end{aligned}$$

The matrix $\begin{pmatrix} 1 \\ d'_3 \\ \vdots \\ d'_N \end{pmatrix} (1, d'_3, \dots, d'_N)$ is symmetric and its eigenvalues are 0 and $1 + d_3^2 + \dots + d_i^2$.

Let $Q_i = \sqrt{1 + d_3^2 + \dots + d_i^2}$ and let G be an orthogonal matrix ($G^T G$

is the identity matrix) such that $G^T \begin{pmatrix} 1 \\ d'_3 \\ \vdots \\ d'_N \end{pmatrix} (1, d'_3, \dots, d'_N) G = \begin{pmatrix} Q_i^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$,

that is,

$$G = \begin{pmatrix} \frac{1}{Q_N} & \frac{-d'_3}{Q_3} & \cdots & \frac{-d'_i}{Q_{i-1}Q_i} & \cdots & \frac{-d'_N}{Q_{N-1}Q_N} \\ \frac{d'_3}{Q_N} & \frac{1}{Q_3} & \cdots & \frac{-d'_3d'_i}{Q_{i-1}Q_i} & \cdots & \frac{-d'_3d'_N}{Q_{N-1}Q_N} \\ \frac{d'_4}{Q_N} & 0 & \cdots & \frac{-d'_4d'_i}{Q_{i-1}Q_i} & \cdots & \frac{-d'_4d'_N}{Q_{N-1}Q_N} \\ \vdots & \vdots & & \vdots & & \vdots \\ & & & \frac{-d'_{i-1}d'_i}{Q_{i-1}Q_i} & & \\ & & & \frac{Q_{i-1}}{Q_i} & & \\ & & & 0 & & \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \frac{d'_N}{Q_N} & 0 & \cdots & 0 & \cdots & \frac{Q_{N-1}}{Q_N} \end{pmatrix}.$$

Then we have

$$\begin{aligned} & v_1^2 \begin{pmatrix} 1 \\ d'_3 \\ \vdots \\ d'_N \end{pmatrix} (1, d'_3, \dots, d'_N) + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \\ &= v_1^2 G \begin{pmatrix} 1 + d_3^2 + \cdots + d_N^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} G^\top + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \\ &= G \begin{pmatrix} 1 + v_1^2(1 + d_3^2 + \cdots + d_N^2) & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} G^\top. \end{aligned}$$

Therefore, we can change the variables from $(r_{i2}, r_{i3}, \dots, r_{iN})$ to $(r'_{i2}, r'_{i3}, \dots, r'_{iN})$ by

$$\begin{pmatrix} r'_{i2} \\ r'_{i3} \\ \vdots \\ r'_{iN} \end{pmatrix} = \begin{pmatrix} \sqrt{1 + v_1^2(1 + d_3^2 + \cdots + d_N^2)} & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} G^\top \begin{pmatrix} r_{i2} \\ r_{i3} \\ \vdots \\ r_{iN} \end{pmatrix}.$$

We have

$$(r_{i2}, r_{i3}, \dots, r_{iN}) \left(v_1^2 \begin{pmatrix} 1 \\ d'_3 \\ \vdots \\ d'_N \end{pmatrix} (1, d'_3, \dots, d'_N) + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right) \begin{pmatrix} r_{j2} \\ r_{j3} \\ \vdots \\ r_{jN} \end{pmatrix}$$

$$= \sum_{k=2}^N r'_{ik} r'_{jk}.$$

Q.E.D.

Since $c_j > 0$, let $\bar{a}_{ij} = \sqrt{c_j} a_{ij}$.

We have

$$\begin{aligned} B^{x_{i_1, i_2}} &= \sum_{j=1}^N \log\left(\frac{c_j s_{x, j, 2} - c_j^* s_{x, j, 2}^*}{1 + c_j^* s_{x, j, 2}^*} + 1\right) = \sum_{j=1}^N \log\left(\frac{c_j a_{i_1 j} a_{i_2 j} - c_j^* a_{i_1 j}^* a_{i_2 j}^*}{1 + c_j^* a_{i_1 j}^* a_{i_2 j}^*} + 1\right) \\ &= \sum_{j=1}^N \log\left(\frac{\bar{a}_{i_1 j} \bar{a}_{i_2 j} - \bar{a}_{i_1 j}^* \bar{a}_{i_2 j}^*}{1 + \bar{a}_{i_1 j}^* \bar{a}_{i_2 j}^*} + 1\right). \end{aligned}$$

The following Step 1, 2 and 3 refer to the proof of Case 1 and Case 2.

A.2.1. Step 1.

Assume that $a^* = 0$.

For $x \in \{0, 1\}^M$, let $x^{(s)} \in \{0, 1\}^M$ be $x_i^{(s)} = \begin{cases} 0, & \text{if } i \leq s, \\ x_i, & \text{if } i > s, \end{cases}$

$$\text{and let } a^{(s)} = \begin{pmatrix} a_{11}^{(s)} & a_{12}^{(s)} & \cdots & a_{1H}^{(s)} \\ a_{21}^{(s)} & a_{22}^{(s)} & \cdots & a_{2H}^{(s)} \\ & & \vdots & \\ a_{M1}^{(s)} & a_{M2}^{(s)} & \cdots & a_{MH}^{(s)} \end{pmatrix}.$$

By using a blowing up process together with an inductive method of s , we have the following **Case A** or **Case B**.

Case A

$$\begin{aligned} B^{x_{i_1, i_2}} &= \sum_{j=1}^N \log(c_j s_{x, j, m} + 1) = \sum_{j=1}^N \log(\bar{a}_{i_1 j} \bar{a}_{i_2 j} + 1) \\ &= \begin{cases} u_1^2 u_2^2 \cdots u_{i_1}^2 u_{i_1+1} \cdots u_{i_2} \{f_{i_1 i_2}^{(s)} + a_{i_2 i_1}^{(s)} + u_1^2 B_{(s)}^{x_{i_1 i_2}}\}, & i_1 < i_2 \leq s, \\ u_1^2 u_2^2 \cdots u_{i_1}^2 u_{i_1+1} \cdots u_s \{f_{i_1 i_2}^{(s)} + a_{i_2 i_1}^{(s)} + u_1^2 B_{(s)}^{x_{i_1 i_2}}\}, & i_1 \leq s < i_2, \\ u_1^2 u_2^2 \cdots u_s^2 \{f_{i_1 i_2}^{(s)} + \sum_{k=s+1}^N a_{i_1 k}^{(s)} a_{i_2 k}^{(s)} + u_1^2 B_{(s)}^{x_{i_1 i_2}}\}, & s < i_1 < i_2, \end{cases} \end{aligned} \quad (5)$$

where $f_{i_1 i_2}^{(s)}$ and $B_{(s)}^{x_{i_1 i_2}}$ are $x_{i_1 i_2}^{(s)}$ -type functions of $(a_{kl}^{(s)})_{s+1 \leq k \leq M, 1 \leq l \leq N}$ with

$$f_{i_1 i_2}^{(s)} \in \left\langle a_{i_1 1}^{(s)}, \dots, a_{i_1, \min\{i_1-1, s\}}^{(s)}, a_{i_2 1}^{(s)}, \dots, a_{i_2, \min\{i_1-1, s\}}^{(s)} \right\rangle.$$

Case B Fix ℓ, ℓ' such as $\ell \leq s < \ell'$. The ideal $\langle B^{x_{i_1, i_2}}, i_1 < i_2 \rangle$ is generated by the following functions:

$$\begin{cases} u_1^2 u_2^2 \cdots u_{i_1}^2 u_{i_1+1} \cdots u_{i_2} \tilde{a}_{i_2 i_1}^{(s)}, & i_1 < i_2 \leq s, \\ u_1^2 u_2^2 \cdots u_{i_1}^2 u_{i_1+1} \cdots u_{s+1} \tilde{a}_{i_2 i_1}^{(s)}, & i_1 \leq s < i_2, i_1 < \ell, \\ u_1^2 u_2^2 \cdots u_{\ell'}^2 u_{\ell'+1} \cdots u_{s+1}. \end{cases} \quad (6)$$

Also we have

$$d\bar{a} = u_1^{MN-1} u_2^{(M-1)N-1} \cdots u_s^{(M-s+1)N-1} u_{s+1}^{(M-s)N-1} du_1 \cdots du_s.$$

Construct the blow-up of B^x along the submanifold $\{\bar{a}_{ij} = 0, 1 \leq i \leq M, 1 \leq j \leq N\}$.

Let $\bar{a}_{11} = u_1, \bar{a}_{ij} = u_1 \bar{a}'_{ij}, (i, j) \neq (1, 1)$.

Let $\bar{a}''_{i1} = \bar{a}'_{i1} + \sum_{j=2}^N \bar{a}'_{1j} \bar{a}'_{ij}$ for $i \geq 2$.

Then we have

$$B^{x_{i_1, i_2}} = \begin{cases} u_1^2 (\bar{a}'_{i_2 1} + \sum_{j=2}^N \bar{a}'_{1j} \bar{a}'_{i_2 j}) + u_1^4 B'^{x_{i_1, i_2}} = u_1^2 \bar{a}''_{i_2 1} + u_1^4 B'^{x_{i_1, i_2}}, & i_1 = 1, i_2 > 1 \\ u_1^2 (\sum_{j=1}^N \bar{a}'_{i_1 j} \bar{a}'_{i_2 j}) + u_1^4 B'^{x_{i_1, i_2}} & 2 \leq i_1 < i_2 \end{cases}$$

Then for $2 \leq i_1 < i_2$,

$$\begin{aligned} \sum_{j=1}^N \bar{a}'_{i_1 j} \bar{a}'_{i_2 j} &= (\bar{a}''_{i_1 1} - \sum_{j=2}^N \bar{a}'_{1j} \bar{a}'_{i_1 j}) (\bar{a}''_{i_2 1} - \sum_{j=2}^N \bar{a}'_{1j} \bar{a}'_{i_2 j}) + \sum_{j=2}^N \bar{a}'_{i_1 j} \bar{a}'_{i_2 j} \\ &= f_{i_1 i_2}^{(1)} + (\bar{a}_{i_1 2}, \dots, \bar{a}_{i_1 N}) \begin{pmatrix} \bar{a}'_{i_2 12} \\ \bar{a}'_{i_2 13} \\ \vdots \\ \bar{a}'_{i_2 1N} \end{pmatrix} (\bar{a}'_{12}, \dots, \bar{a}'_{1N}) \begin{pmatrix} \bar{a}'_{i_2 2} \\ \bar{a}'_{i_2 3} \\ \vdots \\ \bar{a}'_{i_2 N} \end{pmatrix} \\ &\quad + (\bar{a}'_{i_1 2}, \dots, \bar{a}'_{i_1 N}) \begin{pmatrix} \bar{a}'_{i_2 2} \\ \bar{a}'_{i_2 3} \\ \vdots \\ \bar{a}'_{i_2 N} \end{pmatrix}, \end{aligned}$$

where $f_{i_1 i_2}^{(1)} = \bar{a}''_{i_1 1} (\bar{a}''_{i_2 1} - \sum_{j=2}^N \bar{a}'_{1j} \bar{a}'_{i_2 j}) - (\sum_{j=2}^N \bar{a}'_{1j} \bar{a}'_{i_1 j}) \bar{a}''_{i_2 1}$ and $B'^{x_{i_1, i_2}}$ are $x_{i_1 i_2}^{(1)}$ -type functions of

$$\begin{pmatrix} \bar{a}''_{21} & \bar{a}'_{22} & \cdots & \bar{a}'_{2N} \\ \bar{a}''_{31} & \bar{a}'_{32} & \cdots & \bar{a}'_{3N} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{a}''_{M1} & \bar{a}'_{M2} & \cdots & \bar{a}'_{MN} \end{pmatrix} \text{ with } f_{i_1 i_2}^{(1)} \in \langle \bar{a}''_{i_1 1}, \bar{a}''_{i_2 1} \rangle.$$

By Lemma 4, after changing variables, we have

$$(\bar{a}'_{i_1 2}, \bar{a}'_{i_1 3}, \dots, \bar{a}'_{i_1 N}) \left(\begin{pmatrix} \bar{a}'_{i_2 12} \\ \bar{a}'_{i_2 13} \\ \vdots \\ \bar{a}'_{i_2 1N} \end{pmatrix} (\bar{a}'_{12}, \bar{a}'_{13}, \dots, \bar{a}'_{1N}) + \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right) \begin{pmatrix} \bar{a}'_{i_2 2} \\ \bar{a}'_{i_2 3} \\ \vdots \\ \bar{a}'_{i_2 N} \end{pmatrix}$$

$$= \sum_{j=2}^N \bar{a}''_{i_1 j} \bar{a}''_{i_2 j}.$$

Let $a_{ij}^{(1)} = \bar{a}''_{ij}$ for $2 \leq i \leq M, 1 \leq j \leq N$ and $a_{1j}^{(1)} = \bar{a}'_{1j}$ for $2 \leq j \leq N$.

We have **Case A** with $s = 1$, i.e., $B^{x_{i_1 i_2}} = u_1^2(a_{i_2 1}^{(1)} + u_1^2 B_{(1)}^{x_{i_1 i_2}})$ for $i_2 > 1$, $B^{x_{i_1 i_2}} = u_1^2(f_{i_1 i_2}^{(1)} + \sum_{j=2}^N a_{i_1 j}^{(1)} a_{i_2 j}^{(1)} + u_1^2 B_{(1)}^{x_{i_1 i_2}})$ for $1 < i_1 < i_2$. Note that $I = \langle \bar{a}''_{i_1 1}, \bar{a}'_{i_1 1}, \dots, \bar{a}'_{i_H 1} \rangle = \langle a_{i_1 1}^{(1)}, \dots, a_{i_H 1}^{(1)} \rangle$ since the coordinate change is linear. Therefore, we also have $f_{i_1 i_2}^{(1)}$ and

$$B^{x_{i_1 i_2}} \text{ are } x_{i_1 i_2}^{(1)}\text{-type functions of } \begin{pmatrix} a_{21}^{(1)} & a_{22}^{(1)} & \cdots & a_{2N}^{(1)} \\ a_{31}^{(1)} & a_{32}^{(1)} & \cdots & a_{3N}^{(1)} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M1}^{(1)} & a_{M2}^{(1)} & \cdots & a_{MN}^{(1)} \end{pmatrix} \text{ with } f_{i_1 i_2}^{(1)} \in \langle a_{i_1 1}^{(1)}, a_{i_2 1}^{(1)} \rangle.$$

Assume **Case A**. Construct the blow-up of function (5) along the submanifold $\{a_{ij}^{(s)} = 0, s+1 \leq i \leq M, 1 \leq j \leq N\}$.

Let $a_{ij}^{(s)} = u_{s+1} a'_{ij}^{(s)}$ for $s+1 \leq i \leq M, 1 \leq j \leq N$ and $a_{ij}^{(s)} = a'_{ij}^{(s)}$ for $i_1 < i_2 \leq s$.

We have

$$B_{(s)}^{x_{i_1 i_2}} = \begin{cases} u_1^2 u_2^2 \cdots u_{i_1}^2 u_{i_1+1} \cdots u_{i_2} \{f_{i_1 i_2}^{(s)} + a'_{i_2 i_1}^{(s)} + u_1^2 B_{(s)}^{x_{i_1 i_2}}\}, & i_1 < i_2 \leq s, \\ u_1^2 u_2^2 \cdots u_{i_1}^2 u_{i_1+1} \cdots u_s u_{s+1} \{f_{i_1 i_2}^{(s)}/u_{s+1} + a'_{i_2 i_1}^{(s)} + u_1^2 B_{(s)}^{x_{i_1 i_2}}/u_{s+1}\}, & i_1 \leq s < i_2, \\ u_1^2 u_2^2 \cdots u_s^2 u_{s+1}^2 \{f_{i_1 i_2}^{(s)}/u_{s+1}^2 + \sum_{j=s+1}^N a'_{i_1 j}^{(s)} a'_{i_2 j}^{(s)} + u_1^2 B_{(s)}^{x_{i_1 i_2}}/u_{s+1}^2\}, & s < i_1 < i_2. \end{cases}$$

We may set $a'_{s+1, s+1}^{(s)} = 1$ or $a'_{\ell' \ell}^{(s)} = 1$ for $s+1 \leq \ell' \leq M$ and $1 \leq \ell \leq s$.

If $a'_{\ell' \ell}^{(s)} = 1$, we may assume $a'_{i_2 i_1}^{(s)}$ for $1 \leq i_1 < i_2 \leq s$, and for $1 \leq i_1 \leq \ell-1, i_2 > s$ are in a neighborhood of 0. Since $f_{\ell' \ell}^{(s)} \in \langle a_{\ell' 1}^{(s)}, \dots, a_{\ell, \ell-1}^{(s)}, a_{\ell' 1}^{(s)}, \dots, a_{\ell', \ell-1}^{(s)} \rangle$, we have $f_{\ell' \ell}^{(s)}/u_{s+1} + a'_{\ell' \ell}^{(s)} \neq 0$, and $\langle B_{(s)}^{x_{i_1 i_2}} \mid \ell \leq i_1 < i_2, i_2 > s \rangle = \langle u_1^2 u_2^2 \cdots u_{\ell'}^2 u_{\ell+1} \cdots u_s u_{s+1} \rangle$

$$\text{Let } \tilde{a}_{i_2 i_1}^{(s)} = \begin{cases} f_{i_1 i_2}^{(s)} + a'_{i_2 i_1}^{(s)} + u_1^2 B_{(s)}^{x_{i_1 i_2}}, & i_1 < i_2 \leq s, \\ f_{i_1 i_2}^{(s)}/u_{s+1} + a'_{i_2 i_1}^{(s)} + u_1^2 B_{(s)}^{x_{i_1 i_2}}/u_{s+1}, & i_1 < \ell, i_2 > i_1. \end{cases}$$

Jacobian $|(\frac{\partial \tilde{a}_{i_2 i_1}^{(s)}}{\partial a_{i_2 i_1}^{(s)}})|$ is not zero, since $f_{i_1 i_2}^{(s)} \in \langle a_{i_1 1}^{(s)}, \dots, a_{i_1, \min\{i_1-1, s\}}^{(s)}, a_{i_2 1}^{(s)}, \dots, a_{i_2, \min\{i_1-1, s\}}^{(s)} \rangle$.

Therefore, we have **Case B**.

Next assume $a'_{s+1, s+1}^{(s)} = 1$.

Set $a''_{i_2, s+1}^{(s)} = a'_{i_2, s+1}^{(s)} + \sum_{j=s+2}^N a'_{s+1, j}^{(s)} a'_{i_2 j}^{(s)}$ for $i_2 \geq s+2$.

Then for $s+2 \leq i_1 < i_2$,

$$f_{i_1 i_2}^{(s)}/u_{s+1}^2 + \sum_{j=s+1}^N a'_{i_1 j}^{(s)} a'_{i_2 j}^{(s)}$$

$$\begin{aligned}
&= f_{i_1 i_2}^{(s)}/u_{s+1}^2 + (a''_{i_1, s+1} - \sum_{j=s+2}^N a'_{s+1, j} a'_{i_1 j}^{(s)}) (a''_{i_2, s+1} - \sum_{j=s+2}^N a'_{s+1, j} a'_{i_2 j}^{(s)}) + \sum_{j=s+2}^N a'_{i_1 j}^{(s)} a'_{i_2 j}^{(s)} \\
&= f_{i_1 i_2}^{(s)}/u_{s+1}^2 + a''_{i_2, s+1} (a''_{i_1, s+1} - \sum_{j=s+2}^N a'_{s+1, j} a'_{i_1 j}^{(s)}) - a''_{i_1, s+1} (\sum_{j=s+2}^N a'_{s+1, j} a'_{i_2 j}^{(s)}) \\
&\quad + (\sum_{j=s+2}^N a'_{s+1, j} a'_{i_1 j}^{(s)}) (\sum_{j=s+2}^N a'_{s+1, j} a'_{i_2 j}^{(s)}) + \sum_{j=s+2}^N a'_{i_1 j}^{(s)} a'_{i_2 j}^{(s)}.
\end{aligned}$$

$$\text{Let } f_{i_1 i_2}^{(s+1)} = f_{i_1 i_2}^{(s)}/u_{s+1}^2 + a''_{i_2, s+1} (a''_{i_1, s+1} - \sum_{j=s+2}^N a'_{s+1, j} a'_{i_1 j}^{(s)}) - a''_{i_1, s+1} (\sum_{j=s+2}^N a'_{s+1, j} a'_{i_2 j}^{(s)}).$$

Then $f_{i_1 i_2}^{(s+1)}$ is an $x_{i_1, i_2}^{(s+1)}$ -type function of $\begin{pmatrix} a'_{s+2, 1}^{(s)} & a'_{s+2, 2}^{(s)} & \cdots & a'_{s+2, N}^{(s)} \\ \vdots & \vdots & \vdots & \vdots \\ a'_{M1}^{(s)} & a'_{M2}^{(s)} & \cdots & a'_{MN}^{(s)} \end{pmatrix}$ with

$$f_{i_1 i_2}^{(s+1)} \in \langle a'_{i_1 1}^{(s)}, \dots, a''_{i_1, s+1}, a'_{i_2 1}^{(s)}, \dots, a''_{i_2, s+1} \rangle.$$

We have, by Lemma 4,

$$\begin{aligned}
&(\sum_{j=s+2}^N a'_{s+1, j} a'_{i_1 j}^{(s)}) (\sum_{j=s+2}^N a'_{s+1, j} a'_{i_2 j}^{(s)}) + \sum_{j=s+2}^N a'_{i_1 j}^{(s)} a'_{i_2 j}^{(s)} \\
&= \sum_{j=s+2}^N a''_{i_1 j} a''_{i_2 j}.
\end{aligned}$$

Let $a_{i_2 i_1}^{(s+1)} = a'_{i_2 i_1}^{(s)}$ for $i_1 \leq s$, $i_1 < i_2$ and $a_{i_2 i_1}^{(s+1)} = a''_{i_2 i_1}^{(s)}$ for $s < i_1 < i_2$ and $(i_1, i_2) \neq (s+1, s+1)$. Also let $f_{i_1 i_2}^{(s+1)} = f_{i_1 i_2}^{(s)}$ for $i_1 < i_2 \leq s$, $f_{i_1 i_2}^{(s+1)} = f_{i_1 i_2}^{(s)}/u_{s+1}$ for $i_1 \leq s < i_2$, $f_{s+1, i_2}^{(s+1)} = f_{s+1, i_2}^{(s)}/u_{s+1}^2$ for $i_2 > s+1$, $B_{(s+1)}^{x_{i_1 i_2}} = B_{(s)}^{x_{i_1 i_2}}$ for $i_1 < i_2 \leq s$, $B_{(s+1)}^{x_{i_1 i_2}} = B_{(s)}^{x_{i_1 i_2}}/u_{s+1}$ for $i_1 \leq s < i_2$, $B_{(s+1)}^{x_{i_1 i_2}} = B_{(s)}^{x_{i_1 i_2}}/u_{s+1}^2$ for $s < i_1 < i_2$.

Then we have **Case A** with $s+1$.

A.2.2. Step 2.

In Step 1, we finally have **Case B** type. By blowing up the function in **Case B** again, we may obtain the log canonical threshold. However, since we have the same inductive results for

$$\left\langle \sum_{j=1}^N a_{i_1 j} a_{i_2 j} : 1 \leq i_1 < i_2 \leq M \right\rangle \quad (7)$$

instead of

$$\langle B^{x_{i_1 i_2}} : 1 \leq i_1 < i_2 \leq M \rangle$$

if $s \geq \min\{M, N\}$ and thus both log canonical thresholds are the same, we can obtain the log canonical threshold much more simply by using Eq. (7).

Now we consider the log canonical threshold of $\sum_{1 \leq i_1 < i_2 \leq M} (\sum_{j=1}^N a_{i_1 j} a_{i_2 j})^2$.

In Step 2, we use the same symbol a rather than $a^{(\ell)}$ for the sake of simplicity.

We need to consider the ideal generated by the following function with the inductive method with s .

$$\begin{cases} u_1^2 u_2^2 \cdots u_s^2 a_{i_2 i_1}, & i_1 \leq s, i_1 < i_2 \leq M, \\ u_1^2 u_2^2 \cdots u_s^2 \sum_{j=s+1}^N a_{i_1 j} a_{i_2 j} & s+1 \leq i_1 < i_2 \leq M, \end{cases} \quad (8)$$

and

$$da = \prod_{j=1}^s u_j^{(M-j+1)(N-j+1)+(2M-j)(j-1)-1} du_1 du_2 \cdots du_s.$$

Construct the blow-up of $\sum_{j=1}^N a_{i_1 j} a_{i_2 j}$ along the submanifold $\{a_{ij} = 0, 1 \leq i \leq M, 1 \leq j \leq N\}$. Then we have by Lemma 4

$$\begin{cases} u_1^2 a_{i_2 1}, & 1 < i_2 \leq M, \\ u_1^2 (f_{i_1 i_2}^{(1)} + \sum_{j=2}^N a_{i_1 j} a_{i_2 j}) & 2 \leq i_1 < i_2 \leq M, \end{cases}$$

and $da = u_1^{MN-1} du_1 du_2 \cdots du_N$.

By using Lemma 1 again, we have Eq. (8) with $s = 1$ as

$$\begin{cases} u_1^2 a_{i_2 1}, & 1 < i_2 \leq M, \\ u_1^2 (\sum_{j=2}^N a_{i_1 j} a_{i_2 j}) & 2 \leq i_1 < i_2 \leq M. \end{cases}$$

Assume the ideal generated by Eq. (8). Construct the blow-up of function (8) along the submanifold $\{a_{i_2 i_1} = 0, 1 \leq i_1 < i_2 \leq M, i_1 \leq s, a_{i_3 i_4} = 0, s+1 \leq i_3 \leq M, s+1 \leq i_4 \leq N\}$.

Then we have

$$\begin{cases} u_1^2 \cdots u_s^2 u_{s+1} a_{i_2 i_1}, & 1 \leq i_1 < i_2 \leq M, i_1 \leq s, \\ u_1^2 \cdots u_s^2 u_{s+1}^2 \sum_{j=s+1}^N a_{i_1 j} a_{i_2 j}, & s+1 \leq i_1 < i_2 \leq M, \end{cases}$$

and

$$u_{s+1}^{(M-s)(N-s)+(2M-1-s)s/2-1} \prod_{j=1}^s u_j^{(M-j+1)(N-j+1)+(2M-j)(j-1)-1} du_1 du_2 \cdots du_s du_{s+1},$$

where we can set $a_{21} = 1$ or $a_{s+1, s+1} = 1$.

If $a_{21} = 1$, we have

$$\frac{(M-j+1)(N-j+1) + (2M-j)(j-1)}{4}, j = 1, \dots, s$$

and

$$\frac{(M-s)(N-s) + (2M-1-s)s/2}{2},$$

as candidates for the log canonical threshold.

If $a_{s+1,s+1} = 1$, then setting the variables the same as in Step 1 and using Lemma 1, we have

$$\begin{cases} u_1^2 \cdots u_s^2 u_{s+1} a_{i_2 i_1}, & 1 \leq i_1 < i_2 \leq M, i_1 \leq s, \\ u_1^2 \cdots u_s^2 u_{s+1}^2 a_{i_2, s+1}, & s+1 < i_2 \leq M, \\ u_1^2 \cdots u_s^2 u_{s+1}^2 \sum_{j=s+2}^N a_{i_1 j} a_{i_2 j}, & s+2 \leq i_1 < i_2 \leq M, \end{cases} \quad (9)$$

and

$$da = u_{s+1}^{(M-s)(N-s)+(2M-1-s)s/2-1} \prod_{j=1}^s u_k^{(M-j+1)(N-j+1)+(2M-j)(j-1)-1} d u d a.$$

Construct the blow-up of the ideal generated by Eq. (9) along the submanifold $\{a_{i_2 i_1} = 0, 1 \leq i_1 < i_2 \leq M, i_1 \leq s, u_{s+1} = 0\}$.

Then if $a_{ij} = u_{s+1} a_{ij}$, we have the ideal (8) with $s+1$ as

$$\begin{cases} u_1^2 \cdots u_s^2 u_{s+1}^2 a_{i_2 i_1}, & i_1 \leq s+1, 1 \leq i_1 < i_2 \leq M, \\ u_1^2 \cdots u_s^2 u_{s+1}^2 \sum_{j=s+2}^N a_{i_1 j} a_{i_2 j}, & s+1 \leq i_1 < i_2 \leq M, \end{cases}$$

and

$$da = u_{s+1}^{(M-s)(N-s)+(2M-1-s)s-1} \prod_{k=1}^s u_k^{(M-k+1)(N-k+1)+(2M-k)(k-1)-1} d u d a.$$

If $a_{ij} = a_{21} a_{ij}$, $u_{s+1} = a_{21} u_{s+1}$, we have

$$\begin{cases} u_1^2 \cdots u_s^2 u_{s+1} a_{21}^2 \\ u_1^2 \cdots u_s^2 u_{s+1} a_{21}^2 a_{i_2 i_1}, & 1 \leq i_1 < i_2 \leq M, i_1 \leq s+1, (i_1, i_2) \neq (1, 2), \\ u_1^2 \cdots u_s^2 u_{s+1} a_{21}^2 a_{i_2, s+1}, & s+1 < i_2 \leq M, \\ u_1^2 \cdots u_s^2 u_{s+1} a_{21}^2 \sum_{j=s+2}^N a_{i_1 j} a_{i_2 j}, & s+2 \leq i_1 < i_2 \leq M, \end{cases}$$

and

$$u_{s+1}^{(M-s)(N-s)+(2M-1-s)s/2-1} a_{21}^{(M-s)(N-s)+(2M-1-s)s-1} \prod_{k=1}^s u_k^{(M-k+1)(N-k+1)+(2M-k)(k-1)-1} d u d b,$$

which have

$$\frac{(M-j+1)(N-j+1)+(2M-j)(j-1)}{4}, j = 1, \dots, s+1,$$

and

$$\frac{(M-s)(N-s)+(2M-1-s)s/2}{2}.$$

as candidates for the log canonical threshold.

Finally, we have

$$u_1^2 u_2^2 \cdots u_s^2 a_{i_2, i_1}, \quad 1 \leq i_1 < i_2 \leq M, i_1 \leq s$$

where $s = \min\{N, M\}$ and

$$da = \prod_{k=1}^s u_k^{(M-k+1)(N-k+1)+(2M-k)(k-1)-1} du db,$$

and obtain the poles

$$\frac{(M-k+1)(N-k+1)+(2M-k)(k-1)}{4}, k = 1, \dots, s,$$

and

$$\frac{(2M-1-s)s}{4}.$$

Therefore, we have the log canonical threshold $\frac{MN}{4}$ if $M > N$ and $\frac{M(M-1)}{4}$ if $M \leq N$.

A.2.3. Step 3.

We obtain $\frac{MN}{4}$ as the log canonical threshold of

$$\begin{aligned} & \langle c_j(c_j^{m-2} - c_j^{m-3} + c_j^{m-4} - \dots + (-1)^m) a_{i_1 j} a_{i_2 j} \cdots a_{i_m j} : i_1 < i_2 < \dots < i_m, j = 1, \dots, N \rangle \\ & = \langle a_{i_1 j} a_{i_2 j} : i_1 < i_2, j = 1, \dots, N \rangle. \end{aligned}$$

by using Lemma 2.

By Lemma 3, we have

$$\begin{aligned} & \langle b_i, B^{x_{i_1, i_2}} : 1 \leq i, i_1, i_2, \leq M, i_1 < i_2 \rangle \subset \langle b_i, B^x : 1 \leq i \leq M, |x| \geq 2 \rangle \\ & \subset \langle b_i, c_j(c_j^{m-2} - c_j^{m-3} + \dots + (-1)^m) a_{i_1 j} a_{i_2 j} \cdots a_{i_m j} \\ & \quad : 1 \leq i \leq M, 1 \leq i_1 < i_2 < \dots < i_m \leq M, 1 \leq j \leq N \rangle. \end{aligned}$$

Let λ_0 be the log canonical threshold of the ideal $\langle B^x : |x| \geq 2 \rangle$.

We need the log canonical threshold λ_1 of the ideal $\langle b_i, B^x : 1 \leq i \leq M, |x| \geq 2 \rangle$. By Lemma 2, λ_1 corresponds to $\frac{M}{2} + \lambda_0$.

By using Step 1, 2 and Lemma 1, we have

$$\frac{M}{2} + \frac{MN}{4} \geq \lambda_1 = \frac{M}{2} + \lambda_0 \geq \frac{M}{2} + \begin{cases} \frac{MN}{4}, & M > N, \\ \frac{M(M-1)}{4}, & M \leq N, \end{cases}$$

and therefore we have Theorem 2 Case 1 and Case 2.

A.2.4. The case of $N = 1$.

Assume that $a_{11}^*, a_{21}^*, \dots, a_{r1}^* \neq 0$ and $a_{r+1,1}^*, a_{r+2,1}^*, \dots, a_{M,1}^* = 0$.

If $r = 0$ then we have $\lambda = \frac{M}{2} + \frac{M}{4}$, $\theta = \begin{cases} 2, & M = 2, \\ 1, & M \geq 3, \end{cases}$ as the log canonical threshold of $\sum_{i=1}^M b_i^2 + \sum_{1 \leq i \leq M, 1 \leq i_1 < i_2 \leq M} (c_1 a_{i_1 1} a_{i_2 1})^2$.

Assume that $r > 0$.

By setting

$$\bar{a}_{i1} = \log(1 + c_1 a_{i11} a_{i1}) - \log(1 + c_1^* a_{11}^* a_{i1}^*),$$

for $i = 2, \dots, r$,

$$\bar{a}_{i1} = \log(1 + c_1 a_{i11} a_{i1}),$$

for $i = r + 1, \dots, M$, and we only need to consider the ideal

$$\langle b_1, \dots, b_M, \bar{a}_{21}, \dots, \bar{a}_{M1}, \tilde{B}^x | x \in \mathcal{I} \rangle,$$

$$\begin{aligned} \text{where } \tilde{a}_{i1} &= \begin{cases} a_{i11}, & i = 1, \\ c_1^* a_{11}^* a_{i1}^* / c_1 a_{i11}, & i \neq 1, \end{cases} \quad \tilde{s}_{x,1,m} = \sum_{x_{i_1} = x_{i_2} = \dots = x_{i_m} = 1} \tilde{a}_{i_1 1} \cdots \tilde{a}_{i_m 1}, \text{ and } \tilde{B}^x \\ &= \log\left(\frac{\sum_{m=2}^{|x|} \{c_1(c_1^{m-2} - c_1^{m-3} + c_1^{m-4} - \dots + (-1)^m) \tilde{s}_{x,1,m} - c_1^*(c_1^{*m-2} - \dots + (-1)^m) s_{x,1,m}^*\}}{\sum_{m=2}^{|x|} c_1^*(c_1^{*m-2} - c_1^{*m-3} + c_1^{*m-4} - \dots + (-1)^m) s_{x,1,m}^* + 1} + 1\right). \end{aligned}$$

If $r = 1, 2$ then we have $\lambda = \frac{M}{2} + \frac{M-1}{2}$, $\theta = 1$ by considering the ideal

$$\langle b_1, \dots, b_M, \bar{a}_{21}, \dots, \bar{a}_{M1} \rangle.$$

Assume $r \geq 3$. We need to consider the ideal

$$\begin{aligned} &\left\langle b_1, \dots, b_M, \bar{a}_{21}, \dots, \bar{a}_{M1}, \log\left(c_1 \frac{c_1^{*2} a_{11}^{*2} a_{21}^* a_{31}^*}{c_1^2 a_{11}^2} + 1\right) - \log(c_1^* a_{21}^* a_{31}^* + 1), \right. \\ &\log\left(c_1(c_1 - 1) \frac{a_{11} c_1^{*2} a_{11}^{*2} a_{21}^* a_{31}^*}{c_1^2 a_{11}^2} + c_1^* a_{11}^* a_{21}^* + c_1^* a_{11}^* a_{31}^* + \frac{c_1^{*2} a_{11}^{*2} a_{21}^* a_{31}^*}{c_1 a_{11}^2} + 1\right) \\ &\left. - \log(c_1^*(c_1^* - 1)(a_{11}^* a_{21}^* a_{31}^*) + c_1^* a_{11}^* a_{21}^* + c_1^* a_{11}^* a_{31}^* + c_1^* a_{21}^* a_{31}^* + 1)\right\rangle. \end{aligned}$$

By setting $\bar{c}_1 = \log\left(\frac{c_1^{*2} a_{11}^{*2} a_{21}^* a_{31}^*}{c_1 a_{11}} + 1\right) - \log(c_1^* a_{21}^* a_{31}^* + 1)$, we have

$$\begin{aligned} &\langle b_1, \dots, b_M, \bar{a}_{21}, \dots, \bar{a}_{M1}, \bar{c}_1, \\ &\log\left(\left(\frac{c_1^* a_{11}^{*2}}{a_{11}^2} - 1\right) a_{11} c_1^* a_{21}^* a_{31}^* + c_1^* a_{11}^* a_{21}^* + c_1^* a_{11}^* a_{31}^* + c_1^* a_{21}^* a_{31}^* + 1\right) \\ &- \log(c_1^*(c_1^* - 1)(a_{11}^* a_{21}^* a_{31}^* + c_1^* a_{11}^* a_{21}^* + c_1^* a_{11}^* a_{31}^* + c_1^* a_{21}^* a_{31}^* + 1)) \rangle. \end{aligned}$$

We have $\lambda = \frac{M}{2} + \frac{M+1}{2}$, $\theta = 1$ by setting $\bar{a}_{11} = \log\left(\left(\frac{c_1^* a_{11}^{*2}}{a_{11}^2} - 1\right) a_{11} c_1^* a_{21}^* a_{31}^* + c_1^* a_{11}^* a_{21}^* + c_1^* a_{11}^* a_{31}^* + c_1^* a_{21}^* a_{31}^* + 1\right) - \log(c_1^*(c_1^* - 1)(a_{11}^* a_{21}^* a_{31}^* + c_1^* a_{11}^* a_{21}^* + c_1^* a_{11}^* a_{31}^* + c_1^* a_{21}^* a_{31}^* + 1))$.

By Lemma 3, we have not only Case 4 but also Case 3.

A.2.5. The case of $M = 2$.

We have $\lambda = \frac{3}{2}, \theta = \begin{cases} 2, & N = 1, r = 0, \\ 1, & \text{otherwise,} \end{cases}$ as the log canonical threshold of $b_1^2 + b_2^2 + (\sum_{j=1}^N \log(c_j a_{1j} a_{2j} + 1) - \sum_{j=1}^N \log(c_j^* a_{1j}^* a_{2j}^* + 1))^2$.

A.2.6. Naive Bayesian networks.

If $c_1^{*'} = 0$, then we have $\lambda = \frac{M}{2} + \frac{1}{2}, \theta = \begin{cases} 3, & M = 2, \\ 1, & M \geq 3, \end{cases}$ as the log canonical threshold of

$$\sum_{i=1}^M b_i^2 + \sum_{1 \leq i \leq M, 1 \leq i_1 < i_2 \leq M} (c_1^{*'} a_{i_1 1} a_{i_2 1})^2.$$

If $c_1^{*'} \neq 0$, then we have the same results in subsection A.2.4.